JOURNAL OF MULTIVARIATE ANALYSIS 13, 310–327 (1983)

The Inverse Partial Correlation Function of a Time Series and Its Applications

R. J. BHANSALI

University of Liverpool

Communicated by E. J. Hannan

The concept of the inverse correlation function of a stationary process was introduced by Cleveland (*Technometrics* **14** (1972), 277–293). The inverse partial correlation function of a stationary process may intuitively be thought of as the corresponding extension of the concept of the partial correlation function. A precise mathematical definition of this function is given. Its importance in describing the structure of a moving average of finite order h is discussed. Having observed $X_1,\ldots,X_T$, the autoregressive method of estimating the inverse correlations is employed for constructing sample estimates of the inverse partial correlations. For the hth-order moving average process, the estimates beyond h are, as $T\to\infty$, asymptotically independent normally distributed with 0 mean and variance T^{-1}. Their use for estimating h and for testing hypotheses concerning h is examined.

1. INTRODUCTION

Let x_t $(t=0,\pm1,\ldots)$ be a moving average process of order h

$$x_t=\sum_{j=0}^{h}\beta_h(j)\,\varepsilon_{t-j},\qquad \beta_h(0)=1, \tag{1.1}$$

where h is finite, ε_t is a sequence of uncorrelated random variables with 0 mean, variance σ^2, say, and the $\beta_h(j)$ are real coefficients such that the polynomial

$$B_h(z)=\sum_{j=0}^{h}\beta_h(j)\,z^j$$

Received October 5, 1981; revised April 28, 1982.

Key words: Inverse covariance function, inverse correlation function, partial correlation function, inverse partial correlation function, autoregressive spectral estimate, moving average model.

0047-259X/83 $3.00

is bounded away from 0, $|z| \leqslant 1$. Let $R(u) = Ex_t x_{t+u}$ denote the covariance function, $r(u) = R(u)/R(0)$ the correlation function and

$$f(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} R(u) \exp(-iu\lambda)$$

the spectral density function of x_t. We have, $f(\lambda) \neq 0$ $(-\infty < \lambda < \infty)$ and

$$\sum_{u=-\infty}^{\infty} |R(u)| < \infty.$$

Hence, the quantity

$$fi(\lambda) = (1/4\pi^2)\{f(\lambda)\}^{-1} \tag{1.2}$$

exists for all λ and, as in Parzen [27], it is called the inverse spectral density function, its Fourier coefficient

$$Ri(u) = \int_{-\pi}^{\pi} e^{iu\lambda} fi(\lambda)\, d\lambda$$

the inverse covariance function, and

$$ri(u) = Ri(u)/Ri(0)$$

the inverse correlation function of x_t. As discussed by Bhansali [10], the $\beta_h(j)$'s satisfy the Cleveland–Parzen equations

$$\sum_{j=0}^{h} \beta_h(j)\, ri(u-j) = 0 \qquad (u = 1, 2,...), \tag{1.3}$$

which are similar to the well-known Yule–Walker equations (e.g., Box and Jenkins [13, pp. 55–64]) and, by analogy, the quantity $-\beta_h(h)$ may be called the inverse partial correlation coefficient of order h of x_t. Similarly, let $\beta_s(j)$ $(j = 1,..., s)$ be the solutions of the equations

$$\sum_{j=0}^{s} \beta_s(j)\, ri(u-j) = 0 \qquad (u = 1,..., s;\ s = 1, 2,...), \tag{1.4}$$

where $\beta_s(0) = 1$. Then the quantity

$$\pi i(s) = -\beta_s(s) \qquad (s = 1, 2,...), \tag{1.5}$$

may be called the inverse partial correlation function of x_t; for example, Hipel *et al.* [21] and Chatfield [14] make use of this terminology.

In this paper the intuitive definition of the inverse partial correlation

function given in the last paragraph is made more precise by defining this function geometrically in a Hilbert space setting. The inverse partial correlations are shown to play exactly the same role in describing the structure of a finite order moving average process as is played by the partial correlations for a finite order autoregressive process. The latter are considered by Quenouille [28], Anderson [5, pp. 214–223), Barndorff-Nielsen and Schou [1], Ramsey [29], among others, and occupy a prominent place in the methodology for time series analysis developed by Box and Jenkins [13].

2. Definition of the Inverse Partial Correlation Function

Let $Z = \{0, \pm 1, \pm 2, \ldots\}$, $Z^+ = \{1, 2, \ldots\}$ and suppose that $\{x_t, t \in Z\}$ is an arbitrary, real valued stationary process with an absolutely summable covariance function $R(u)$, and nonvanishing spectral density function $f(\lambda)$. Further, let $fi(\lambda)$, $Ri(u)$, and $ri(u)$, respectively, denote the inverse spectral density function, the inverse covariance function, and the inverse correlation function of x_t. (For convenience, the notation used here is the same as in Section 1, even though the two processes are not necessarily the same.) Denote by H_∞ the Hilbert space spanned by the random variables $\{x_t, t \in Z\}$ with inner product

$$(x_t, x_s)_H = R(s - t) \tag{2.1}$$

and norm

$$\|x_t\|_H = \sqrt{R(0)}. \tag{2.2}$$

Let H_{t-1} denote the closed linear subspace of H_∞ generated by the random variables $\{x_s, s < t; t \in Z\}$, $\hat{x}_t$ denote the projection of x_t on H_{t-1}, and put

$$\varepsilon_t = x_t - \hat{x}_t. \tag{2.3}$$

We have (e.g., Hannan [19, p. 137])

$$\|\varepsilon_t\|_H^2 = 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda)\, d\lambda\right\} = \sigma^2, \tag{2.4}$$

say,

$$(\varepsilon_t, \varepsilon_s)_H = 0 \qquad (t \neq s). \tag{2.5}$$

Hence, when viewed as elements of H_∞, the ε_t $(t \in Z)$ are mutually

orthogonal. For the purpose of defining the inverse partial correlation function, however, it will be convenient to view them as elements of a different Hilbert space M_∞ on which the inner product is given by

$$(\varepsilon_t, \varepsilon_s)_M = Ri(s-t) \tag{2.6}$$

and the norm by

$$\|\varepsilon_t\|_M = \sqrt{Ri(0)}. \tag{2.7}$$

Thus M_∞ consists of all finite linear combinations $\sum c_j \varepsilon_{t_j}$ $(t_j \in Z)$, where the c_j are real constants, as well as limits with respect to norm (2.7) of all Cauchy sequences in M_∞.

Let $L(f)$ denote the Hilbert space generated by complex exponentials $\{e^{it\lambda},\ t \in Z,\ \lambda \in [-\pi, \pi]\}$ with inner product

$$(e^{it\lambda}, e^{is\lambda})_L = \int_{-\pi}^{\pi} e^{i(t-s)\lambda} f(\lambda)\, d\lambda.$$

It is well known that the space H_∞ is congruent to $L(f)$, see, for example, Grenander and Rosenblatt [17, p. 64]). Denote by $L^*(fi)$ another Hilbert space, generated by complex exponentials $\{e^{it\lambda},\ t \in Z,\ \lambda \in [-\pi, \pi]\}$ with inner product

$$(e^{it\lambda}, e^{is\lambda})_{L^*} = \int_{-\pi}^{\pi} e^{i(t-s)\lambda} fi(\lambda)\, d\lambda.$$

Then the arguments of Grenander and Rosenblatt [17] may be repeated to show that $L^*(fi)$ is congruent to M_∞. In particular, the element $\varepsilon_t \in M_\infty$ corresponds to $e^{it\lambda} \in L^*(fi)$.

Consider the subspace $M_{j+1,j+s-1}$ of M_∞ generated by $\{\varepsilon_{j+1}, \ldots, \varepsilon_{j+s-1}\}$, where $s > 1$ and $j \in Z$. Let $\hat{\varepsilon}_j$ and $\hat{\varepsilon}_{j+s}$ be the respective projections of ε_j and ε_{j+s} on $M_{j+1,j+s-1}$ and put $\mu_j = \varepsilon_j - \hat{\varepsilon}_j$, $\mu_{j+s} = \varepsilon_{j+s} - \hat{\varepsilon}_{j+s}$.

DEFINITION. The inverse partial correlation function of x_t is the sequence $\{\pi i(s),\ s \in Z^+\}$ defined by

$$\pi i(1) = ri(1), \tag{2.8}$$

and for $s > 1$

$$\pi i(s) = (\mu_j, \mu_{j+s})_M / \|\mu_j\|_M \|\mu_{j+s}\|_M, \tag{2.9}$$

for all $j \in Z$. ∎

We have, by the Cauchy–Schwarz inequality,

$$|\pi i(s)| < 1. \tag{2.10}$$

Suppose that the $\beta_s(j)$ are the solutions of Eqs. (1.4), and put

$$\sigma i^*(s)^2 = \sum_{j=0}^{s} \beta_s(j)\, ri(j), \tag{2.11}$$

$$\sigma i(s)^2 = Ri(0)\, \sigma i^*(s)^2. \tag{2.12}$$

It is well known that the $\beta_s(j)$ may be obtained recursively from the $\beta_{s-1}(j)$ (e.g., Ramsey [29]). The relevant equations are

$$\beta_s(s) = -\left\{ri(s) + \sum_{j=1}^{s} \beta_{s-1}(j)\, ri(s-j)\right\} \Big/ \sigma i^*(s-1)^2, \tag{2.13}$$

$$\beta_s(j) = \beta_{s-1}(j) + \beta_s(s)\, \beta_{s-1}(s-j) \qquad (j = 1,\ldots, s-1), \tag{2.14}$$

$$\sigma i^*(s)^2 = \sigma i^*(s-1)^2\, [1 - \{\beta_s(s)\}^2] \tag{2.15}$$

which may be started by recognizing that $\beta_1(1) = -ri(1)$ and $\sigma i^*(1)^2 = [1 - \{\beta_1(1)\}^2]$. Since

$$\mu_j = \sum_{t=0}^{s-1} \beta_{s-1}(t)\, \varepsilon_{j+t}, \qquad \mu_{j+s} = \sum_{t=0}^{s-1} \beta_{s-1}(t)\, \varepsilon_{j+s-1-t}$$

it is readily verified that the definition of $\pi i(s)$ given above is consistent with (1.5) and they may be determined from Eqs. (2.13)–(2.15).

For $t \in Z$, let M_{t-1}^* denote the subspace of M_∞ generated by $\{\varepsilon_s, s < t\}$ and let $\hat{\varepsilon}_t$ denote the projection of ε_t on M_{t-1}^*. We have

$$\|\varepsilon_t - \hat{\varepsilon}_t\|_M^2 = 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log fi(\lambda)\, d\lambda\right\} = \sigma i^2, \tag{2.16}$$

say, and on using Eq. (2.4), we get

$$\sigma i^2 = 1/\sigma^2. \tag{2.17}$$

Note that the relationship (2.17) between σ^2 and σi^2 is consistent with a remark of Parzen (1974).

Finally, it should also be stated that the definition of $\pi i(s)$ may equivalently be given in λ language on the space $L^*(fi)$. The details, however, are omitted in the interests of space.

3. Properties of the Inverse Partial Correlation Function

A principal application of the inverse partial correlation function is in characterizing the behaviour of a finite order moving average process. The following result, the necessity of which follows immediately from Eqs. (1.3), shows that $\pi i(s)$ may be used for identifying a finite order moving average process.

THEOREM 3.1. *A stationary time series x_t with absolutely summable covariance function and nonvanishing spectral density function is a moving average process of order h if and only if its inverse partial correlation function is zero beyond h.*

Proof of Sufficiency. Since $\pi i(s) = 0$, $s > h$, Eq. (2.14) shows that for all $s > h$, $\beta_s(j) = \beta_h(j)$ $(j = 1,..., h)$, $\beta_s(j) = 0$ $(j = h+1,..., s)$. Thus the $\beta_h(j)$ satisfy the following equation:

$$\int_{-\pi}^{\pi} e^{-iu\lambda} B_h(e^{-i\lambda})\, fi(\lambda)\, d\lambda = 0, \qquad u = -1, -2,...,$$
$$= \sigma i^2, \qquad u = 0,$$

where $B_h(z) \neq 0$, $|z| \leqslant 1$ is the characteristic polynomial of the $\beta_h(j)$. This last result, however, implies that if $v(u)$ $(u = 0, 1,...)$ denote the nonvanishing Fourier coefficients of the function $\eta(\lambda) = B_h(e^{-i\lambda})\, fi(\lambda)$, then the function

$$\kappa(z) = \sum_{u=0}^{\infty} v(u)\, z^u = \sigma i^2 \left[1 + \sum_{u=1}^{\infty} v^*(u)\, z^u\right],$$

where $v^*(u) = v(u)/v(0)$, belongs to the Hardy space H^1 of complex functions, e.g., Rudin [31, p. 369], and $\eta(\lambda) = (2\pi)^{-1}\, \kappa(e^{i\lambda})$. Therefore, on using relations (1.2), (2.17) and results of Szegö [35, pp. 268–269] we may deduce that

$$f(\lambda) = \frac{\sigma^2}{2\pi} \{B_h(e^{-i\lambda})\} \left\{1 + \sum_{u=1}^{\infty} v^*(u)\, e^{iu\lambda}\right\}^{-1},$$

$$\left\{1 + \sum_{u=1}^{\infty} v^*(u)\, e^{iu\lambda}\right\}^{-1} = B_h(e^{i\lambda}),$$

and

$$R(u) = 0, \qquad |u| > h. \quad \blacksquare$$

Suppose now that x_t is a moving average process of order h satisfying Eq. (1.1) and let $\mathbf{b}_h = [\beta_h(1),..., \beta_h(h)]'$ and $\boldsymbol{\rho}_h = [\pi i(1),..., \pi i(h)]'$ be $h \times 1$ vectors. Also let $\boldsymbol{\Gamma}$ denote the mapping, defined by Eqs. (2.13)–(2.15), which transforms $\mathbf{b}_h$ to $\boldsymbol{\rho}_h$. Theorem 3.2 below may be established by using the

results of Barndorff-Nielsen and Schou [7] who earlier established a similar theorem for an autoregressive process. Proof of this theorem is not given, since it is almost identical to that given by these authors.

THEOREM 3.2. *The mapping* Γ *which transforms* $\mathbf{b}_h$ *to* $\boldsymbol{\rho}_h$ *is one-to-one and onto* $(-1, 1)^h$. *Furthermore, both* Γ *and its inverse* Γ^{-1} *are continuously differentiable.*

An implication of Theorem 3.2 is that the class of finite order moving average processes may be smoothly parametrized in terms of the $\pi i(s)$ in exactly the same way as the class of finite order autoregressive processes can be in terms of their partial correlation function. Barndorff-Nielsen and Schou [7] discuss advantages of the latter, which also carry through to moving average processes; see also Akaike [4].

4. ESTIMATION OF THE INVERSE PARTIAL CORRELATION FUNCTION

Let $X_1,..., X_T$ denote the observed time series. A natural method of estimating $\pi i(s)$ is to solve Eqs. (2.13)–(2.15) but with $ri(u)$ replaced by its corresponding sample estimate $\tilde{r}i(u)$, say. Several different methods may be employed for obtaining $\tilde{r}i(u)$ (see Bhansali [10]). Of these, only the autoregressive method will be examined here. It is clear that the window method of estimating $ri(u)$ may also be used for estimating the $\pi i(s)$, and under appropriate regularity conditions, results similar to those given in Theorems 4.1, 4.2, and 5.1 also hold for this method. To save space, however, this method is not considered here. Also, in the simulation experiments reported in Section 6, this method gave discouraging results.

The estimate of $ri(u)$ given by the autoregressive method is of the form $\tilde{r}i(u) = \hat{r}i_k(u)$, where

$$\hat{r}i_k(u) = \hat{R}i_k(u)/\hat{R}i_k(0), \tag{4.1}$$

$$\hat{R}i_k(u) = (2\pi Q)^{-1} \sum_{j=0}^{Q-1} \left\{\hat{f}_k\left(\frac{2\pi j}{Q}\right)\right\}^{-1} \exp\left(iu\frac{2\pi j}{Q}\right) \quad (u = 0, 1,..., k), \tag{4.2}$$

$$\hat{f}_k(\lambda) = (2\pi)^{-1} S_k^2 |1 + \hat{a}_k(1)\exp(-i\lambda) + \cdots + \hat{a}_k(k)\exp(-ik\lambda)|^{-2},$$

$c_1 = \hat{a}_k(1),..., c_k = \hat{a}_k(k)$ are the kth order least-squares estimates of autoregressive coefficients, obtained by minimizing

$$(T-k)^{-1} \sum_{j=0}^{T-1-k} (X_{k+j+1} + c_1 X_{k+j} + \cdots + c_k X_{1+j})^2,$$

with minimum S_k^2 and $Q \geqslant 2k$.

Denote by $\hat{\beta}_{k,s}(j)$ the sth order estimates of the moving average parameters obtained by solving Eqs. (2.13)–(2.15) but with $ri(u)$ replaced by $\hat{ri}_k(u)$. Then

$$\hat{\pi i}_k(s) = -\hat{\beta}_{k,s}(s) \qquad (s = 1, 2, \ldots, k) \tag{4.3}$$

gives the corresponding autoregressive estimate of $\pi i(s)$, and

$$\hat{\sigma i}_k(s)^2 = \hat{\sigma i}_k(s-1)^2[1 - \{\hat{\beta}_{k,s}(s)\}^2] \tag{4.4}$$

that of $\sigma i(s)^2$, where $\hat{\sigma i}_k(0)^2 = \hat{R i}_k(0)$.

As noted by Bhansali [10] the above autoregressive method of obtaining the $\hat{\beta}_{k,s}(j)$ is closely connected with the well-known Durbin's [16] method of estimating the parameters $\beta_h(j)$, of the moving average model (1.1). Its main advantage is computational economy. Closed form estimates of the parameters of moving average models of successively higher orders may be obtained in one pass. Second, for each fitted order, the estimates are guaranteed to be invertible, which is not the case with the maximum likelihood methods, such as those of Hannan [18] and Box and Jenkins [13], of estimating the moving average parameters; see Nicholls [25] and Osborn [26]. In the simulation studies of McClave [22, 23], Durbin's method does not compare favourably with the maximum likelihood methods. These comparisons, however, are based on an unrealistic assumption that the order of the moving average model is known. Thus a two-step procedure recommended by Bhansali [10] may be employed; to save space, the details of this procedure are not repeated here.

The asymptotic distribution of the $\hat{\pi i}_k(s)$ takes a particularly simple form when x_t is a finite order moving average process and $s \geqslant h$. It is only this case that is considered below, since otherwise their asymptotic covariance structure is rather complicated. We suppose that x_t satisfies

ASSUMPTION 1. The process x_t follows Eq. (1.1), $B_h(z) \neq 0$, $|z| \leqslant 1$ and ε_t is a sequence of independent identically distributed random variables with 0 mean, variance σ^2, and finite fourth moment $E(\varepsilon_t^4)$.

Let $\mathbf{b}_s = [\beta_s(1), \ldots, \beta_s(s)]'$, $\boldsymbol{\rho}_s = [\pi i(1), \ldots, \pi i(s)]'$, $\hat{\mathbf{b}}_{k,s} = [\hat{\beta}_{k,s}(1), \ldots, \hat{\beta}_{k,s}(s)]'$ be $s \times 1$ vectors and let $\mathbf{0}_s$ denote an s-dimensional vector of zeroes. When $s \geqslant h$, x_t satisfies Assumption 1 and k is a function of T such that as $T \to \infty$, $k \to \infty$ but the regularity conditions stated in Theorem 4.1 hold, Bhansali [11] has shown that, as $T \to \infty$, $\sqrt{T}\ \{\hat{\mathbf{b}}_{k,s} - \mathbf{b}_s\}$ is asymptotically normally distributed with mean $\mathbf{0}_s$ and covariance matrix $\sigma i^2 \boldsymbol{\Phi}(s)^{-1}$, where σi^2 is given by (2.17) and $\boldsymbol{\Phi}(s)$ has $Ri(u-v)$ in its uth row and vth column. Hence, by a result of Rao [30, p. 388] and our Theorem 3.2, $\sqrt{T}\ \{\hat{\boldsymbol{\rho}}_{k,s} - \boldsymbol{\rho}_s\}$ is

asymptotically normally distributed with mean $\mathbf{0}_s$ and covariance matrix $\mathbf{\Sigma}(s)^{-1}$, where

$$\mathbf{\Sigma}(s) = \frac{\partial \mathbf{\Gamma}^{-1}}{\partial \boldsymbol{\rho}_s^*} \tilde{\mathbf{\Phi}}(s) \frac{\partial (\mathbf{\Gamma}^{-1})^*}{\partial \boldsymbol{\rho}_s^*} \Bigg|_{\pi i(h+1) = \cdots = \pi i(s) = 0};$$

$\boldsymbol{\rho}_s^*$ and $(\mathbf{\Gamma}^{-1})^*$ denote the transposes of $\boldsymbol{\rho}_s$ and $\mathbf{\Gamma}^{-1}$, respectively, and the matrix $\tilde{\mathbf{\Phi}}(s)$ is obtained from $\sigma^2 \mathbf{\Phi}(s)$ by replacing the $\beta_v(j)$ and $\sigma i(v)^2$ by equivalent expressions involving the $\pi i(j)$.

In Theorems 4.1 and 4.2, the results of Barndorff-Nielsen and Schou [7] are used to show that $\hat{\pi} i_k(h)$ is asymptotically independently distributed of $\hat{\pi} i_k(1), \ldots, \hat{\pi} i_k(h-1)$, that $\{\hat{\pi} i_k(h+1), \ldots, \hat{\pi} i_k(s)\}$ are asymptotically independent and identically normally distributed with the same variance and that they are also independently distributed of $\{\hat{\pi} i_k(1), \ldots, \hat{\pi} i_k(h)\}$. We remark that an alternative proof of Theorem 4.2 may also be constructed along the lines of Anderson [5, pp. 217–221].

THEOREM 4.1. *Suppose that Assumption* 1 *holds. Further suppose that the choice of* k *in terms of* T *is such that as* $T \to \infty$, $k \to \infty$ *but* $k^3/T \to 0$ *and*

$$T^{1/2}\{|a(k+1)| + |a(k+2)| + \cdots\} \to 0.$$

Then, as $T \to \infty$, $\sqrt{T}\{\hat{\boldsymbol{\rho}}_{k,h} - \boldsymbol{\rho}_h\}$ *is asymptotically normally distributed with mean* $\mathbf{0}_h$ *and covariance matrix* $\mathbf{\Sigma}(h)^{-1}$, *where* $\mathbf{\Sigma}(h)$ *is of the form*

$$\mathbf{\Sigma}(h) = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0}_{h-1}^* \\ \hline \mathbf{0}_{h-1} & \{1 - \pi i(h)^2\}^{-1} \end{array}\right] \tag{4.5}$$

and $\mathbf{A}_{11} = \mathbf{A}_{11}(\boldsymbol{\rho}_h)$ *denotes an* $(h-1) \times (h-1)$ *matrix.*

THEOREM 4.2. *Under the conditions stated in Theorem* 4.1, *for* $s > h$, $\sqrt{T}\{\hat{\boldsymbol{\rho}}_{k,s} - \boldsymbol{\rho}_s\}$ *is asymptotically normally distributed with mean* $\mathbf{0}_s$ *and covariance matrix* $\mathbf{\Sigma}(s)^{-1}$, *where* $\mathbf{\Sigma}(s)$ *is of the form*

$$\mathbf{\Sigma}(s) = \begin{bmatrix} \mathbf{\Sigma}(h) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{4.6}$$

$\mathbf{I}$ denotes the identity matrix of order $s-h$ and $\mathbf{0}$ stands for a matrix consisting of zeroes.

5. ORDER SELECTION FOR A MOVING AVERAGE PROCESS

There are three different ways in which the estimated inverse partial correlations may be used for selecting the order of a moving average to fit:

(i) The $\hat{\pi}i_k(s)$ may be plotted against s in much the same way as the estimated partial correlations are for determining the order of an autoregressive process; see Hipel *et al.* [21] and Chatfield [14]. An alternative is to plot $\hat{\sigma}i_k(s)^2$ against s; see Whittle [36, p. 37]. In our view, this informal inspection of the data is often the most important part of an entire analysis. Nevertheless, it cannot ordinarily stand by itself, since one usually needs to provide various standards of comparison against which the observed discrepancies can be measured.

(ii) The null hypothesis that the order is h_0 against the alternative that it is $h_0 + h_1$, $h_1 \geqslant 1$, may be tested by forming the statistic

$$Q_1 = T \sum_{s=h_0+1}^{h_0+h_1} \hat{\pi}i_k(s)^2.$$

It follows from Theorem 4.2 that under the null hypothesis Q_1 is asymptotically χ^2 distributed with h_1 degrees of freedom. Thus, this procedure is similar to that suggested by Quenouille [28] for testing the corresponding hypotheses concerning an autoregressive process. It is suitable to use when h_0 and h_1 have been prescribed a priori. Otherwise, it will need to be applied repeatedly, and there is then the problem of determining an appropriate level of significance to use (see Akaike [3]).

(iii) On the assumption that an upper bound H, say, for h is known a priori an estimate $\tilde{h}_k(\alpha)$, say, of h may be constructed by minimising the criterion

$$\text{AICA}_\alpha(s) = T \log \hat{\sigma}i_k(s)^2 + \alpha s \qquad (s = 0, 1, \ldots, H), \tag{5.1}$$

where $\alpha > 1$ is an arbitrary constant. The motivation for using this criterion comes from the work of Akaike [1, 4], Shibata [33], and Bhansali and Downham [12]. A choice of $\alpha = 2$ corresponds to using Akaike's [1] information criterion, while a choice of $\alpha > 2 \log\log T$ and $\alpha = \log T$ produces criteria equivalent to those considered by Hannan [20] and Schwarz [32].

Let

$$\text{FPEA}_\alpha(s) = \hat{\sigma}i_k(s)^2 \, (1 + \alpha s/T). \tag{5.2}$$

If terms which are $O(T^{-2})$ are ignored, then $\text{AICA}_\alpha(s) = \log\{\text{FPEA}_\alpha(s)\}$. Hence, the use of the criterion (5.2) is asymptotically equivalent to that of (5.1).

The asymptotic distribution of $\tilde{h}_k(\alpha)$, as $T \to \infty$, is given below in Theorem 5.1 for a fixed α. It may be established by using Theorem 4.2 of this paper, Theorem 3.1 of Bhansali [11], and an argument similar to that of Shibata [33]. A detailed proof of this theorem is not given since it is analogous to that given by Shibata [33].

THEOREM 5.1. *Under the conditions stated in Theorem* 4.1,

$$\lim_{T\to\infty} Pr\{\tilde{h}_k(\alpha)=j\} = g_\alpha(j),$$

where

$$g_\alpha(j) = p_\alpha(j-h)\,q_\alpha(H-j), \qquad h \leqslant j \leqslant H,$$
$$= 0, \qquad \textit{otherwise},$$
$$p_\alpha(n) = Pr\left\{\bigcap_{j=1}^{n}(Y_j>0)\right\}, \qquad q_\alpha(n) = Pr\left\{\bigcap_{j=1}^{n}(Y_j\leqslant 0)\right\}, \quad n\geqslant 1,$$
$$p_\alpha(0) = q_\alpha(0) = 1, \qquad Y_j = \{(\chi_1^2(1)-\alpha)+\cdots+(\chi_1^2(j)-\alpha)\}$$

and $\{\chi_1^2(i)\}$ *denotes a sequence of independent* χ_1^2 *random variables.*

It is appropriate to compare the procedure described above of estimating h with some of its alternatives:

(i) The first alternative, considered by Bhansali [11], is to estimate h by $\hat{h}_k(\alpha)$, where $\hat{h}_k(\alpha)$ is the value of s for which the criterion

$$\text{FPER}_\alpha(s) = \hat{\sigma}_k(s)^2\,(1+\alpha s/T) \qquad (s=0,\,1,\ldots,\,H) \tag{5.3}$$

attains its minimum value. Here

$$\hat{\sigma}_k(s)^2 = \frac{2\pi}{T}\sum_{j=0}^{T-1} I^{(T)}(\omega_j)\,|\hat{B}_{k,s}(\omega_j)|^{-2}, \qquad \omega_j = 2\pi j/T, \tag{5.4}$$

$$\hat{B}_{k,s}(\lambda) = \sum_{j=0}^{S}\hat{\beta}_{k,s}(j)\exp(-ij\lambda), \qquad I^{(T)}(\lambda) = (2\pi T)^{-1}\left|\sum_{t=1}^{T} X_t\exp(-it\lambda)\right|^2, \tag{5.5}$$

and, as before, $\alpha > 1$ is an arbitrary constant.

(ii) This alternative, examined by Hannan [20], provides three different estimates of h by minimising three different criteria, all of which may be obtained from a criterion of the general form

$$\text{AICM}_\alpha(s) = \log\sigma^*(s)^2 + \alpha s/T \qquad (s=0,\,1,\ldots,\,H), \tag{5.6}$$

by setting $\alpha = 2$, $\alpha > 2\log\log T$ and $\alpha = \log T$, respectively. In (5.6), $\sigma^*(s)^2$ is of the same form as $\hat{\sigma}_k(s)^2$ but with $\hat{B}_{k,s}(\lambda)$ replaced by the transfer function $G_s^*(\lambda)$, say, of the sth order full maximum likelihood estimates of the moving average parameters.

(iii) The last alternative, proposed by Akaike [2], estimates h by minimising a criterion of the same form as (5.6), but with $\alpha = 2$ and $\sigma^*(s)^2$

calculated by using an approximate maximum likelihood procedure analogous to that of Hannan [18].

We note that for each fixed α, $\hat{h}_k(\alpha)$ and $\tilde{h}_k(\alpha)$ have the same asymptotic distribution, and are thus asymptotically equivalent. These in their turn are asymptotically equivalent to the estimate obtained by minimising the criterion (5.6). As compared with both these procedures, however, the estimation of h by minimising the criterion (5.2) or (5.3), has the advantage of computational economy. The first is computationally more expensive because the computation of $\hat{\sigma}_k(s)^2$ is more costly than that of $\hat{\sigma}i_k(s)^2$. Indeed, the latter may be computed at the same time as the $\hat{\beta}_{k,s}(j)$ are, cf. Eq. (4.4). The second because it employs an iterative full maximum likelihood procedure for estimating the moving average parameters for each $s = 1,..., H$.

For applying the FPEA_α, or the FPER_α, criterion there is also the question of how to choose k. But this problem need not be insurmountable. The numerical results given by Bhansali [11] for the FPER_α criterion suggest that although the order selected could change as k is varied, the frequency of this occurring is not unduly high.

A related reference is McClave [24], the main idea behind whose method of estimating h is similar to that suggested here. The actual method of estimating $ri(u)$ and h, however, is different and more complicated. Thus his results do not yield our Theorem 5.1; also see Bhansali [11].

6. Simulation Results

In practice T is finite. The finite sample distribution of the AICA_α and the FPEA_α criteria, however, is difficult to derive analytically. To examine the usefulness of the asymptotic distribution of $\tilde{h}_k(\alpha)$ given in Theorem 5.1 as an approximation to its finite sample distribution, and to compare the relative performance of the FPEA_α criterion with that of the FPER_α and AICM_α criteria, a Monte Carlo investigation was carried out.

A stretch of T observations following the first order moving average process $x_t = \varepsilon_t + \beta\varepsilon_{t-1}$ was generated, with $\beta = \pm 0.3$, ± 0.8 and $T = 48$, 96, 200, 300, and 500. Here ε_t is a computer-produced sequence of (pseudo) normal deviates with 0 mean and variance 1; its method of generation being exactly the same as that described by Bhansali and Downham [12]. Three different values of α were used, namely, $\alpha = 2$, 3, and 4.

The joint frequency distribution of the order selected by the FPEA_α and the FPER_α criteria was determined for each (T, β, α) configuration; the total number of simulations for each such configuration being 100. The $\hat{R}i_k(u)$ and $\hat{r}i_k(u)$ were computed by ysing formulae (4.2) and (4.1), respectively, where, as discussed by Bhansali [10] k was specified arbitrarily. The $\hat{\beta}_{k,s}(j)$ and

$\hat{\pi}i_k(s)$ $(j = 1,..., s;\ s = 1,..., H)$ were then computed in the manner described in Section 4. For both these criteria we set $H = k$.

The frequency distribution of the order selected by the AICM_α criterion, (5.6), was also determined for each (T, β, α) configuration, by setting $H = 10$. The moving average parameters were estimated by an iterative approximate maximum likelihood procedure suggested by Hannan [18], but the modification introduced by Nicholls [25] for ensuring that the estimates are invertible was also incorporated. For deciding whether the iterative procedure has converged, a tolerance level of 10^{-4} was set. Thus if at the nth and $(n + 1)$th iterations, the estimates differ only in the fifth decimal place, then the iterative procedure was deemed to have converged at the nth iteration $(n \geqslant 1)$. Otherwise, the iterations were continued until convergence took place, but with an upper limit of 30 on the total number of iterations, that is, if the convergence was not attained before the number of iterations exceeded 30, the procedure was declared as not having converged.

The observed frequency distributions of the order selected are shown in Table I along with the asymptotic. Strictly speaking, the latter apply only to the FPEA_α and FPER_α criteria, since they have been calculated from Theorem 5.1 by taking $H = k$ and $h = 1$. The asymptotic frequencies change very slowly with H, however, and those shown for $T = 96$, $k = 12$ apply also for the AICM_α criterion. To save space, only the marginal distributions of the order selected by the FPEA_α, FPER_α and AICM_α criteria, and only the results for $\beta = -0.8$, 0.3 and $\alpha = 2$, 4 are shown.

If $T \geqslant 200$, then the observed frequencies of the order selected by all three criteria are close to the asymptotic, except when the FPER_α criterion is applied with $\alpha = 2$ and $T = 200$ and $\beta = 0.3$. The observed frequency of selecting a first order model is then somewhat higher than the asymptotic.

If, on the other hand, $T = 48$ and 96 and $\beta = 0.3$ the frequency of selecting order zero is appreciable for all three criteria and this frequency increases as α increases from 2 to 4. We note that the disparity between the asymptotic and the finite sample behaviour of these three criteria is qualitatively similar to that reported by Bhansali and Downham [12] for estimating the order of an autoregressive process.

On comparing the frequency distributions of the order selected by the AICM_α criterion with those of the FPEA_α and the FPER_α criteria, it may be gleaned that from the point of view of estimating the order of a moving average process, no significant loss in efficiency results from employing either of the latter two criteria. On the other hand, when using the AICM_α criterion, two major computational difficulties were encountered: First, the computer time needed was more than 200 times greater than that required for using either the FPEA_α or the FPER_α criteria. Second, the Hannan [18] iterative procedure for estimating the moving average parameters very often failed to converge for some or all of the fitted orders. This occurred

TABLE I

The Frequency of Order Selected (OS) by the $FPEA_\alpha$, $FPER_\alpha$, and $AICM_\alpha$ Criteria in 100 Realizations of the Process $x_t = \varepsilon_t + \beta\varepsilon_{t-1}$

		$\alpha = 2$							$\alpha = 4$						
	Criterion / OS	$\beta = -0.8$			$\beta = 0.3$			Asymptotic	$\beta = -0.8$			$\beta = 0.3$			Asymptotic
		FPEA	FPER	AICM	FPEA	FPER	AICM		FPEA	FPER	AICM	FPEA	FPER	AICM	
$T = 48$,	0	0	0	0	26	12	12	0	0	0	0	41	33	47	0
$k = 6$	1	74	76	70	54	62	66	73	96	90	80	54	61	46	94
	2	13	11	14	14	11	9	12	3	6	9	4	5	5	4
	>2	13	13	16	6	15	13	15	1	4	11	1	1	2	2
$T = 96$	0	0	0	0	4	3	3	0	0	0	0	12	10	9	0
$k = 12$	1	64	71	75	60	61	57	72	90	94	93	77	82	80	94
	2	16	9	11	17	13	21	11	5	6	3	7	4	9	4
	3–6	18	15	11	17	17	17	13	5	0	4	4	4	2	2
	>6	2	5	3	2	6	2	4	0	0	0	0	0	0	0
$T = 200$	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0
$k = 25$	1	73	76	75	67	82	67	71	93	98	96	91	96	96	94
	2	6	11	11	11	6	9	11	5	2	4	5	3	3	4
	3–6	19	11	13	18	8	15	13	2	0	0	3	0	1	2
	>6	2	2	1	3	4	9	5	0	0	0	0	0	0	0
$T = 300$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$k = 25$	1	65	74	69	68	74	70	71	90	92	91	94	97	96	94
	2	12	11	15	12	9	12	11	6	6	7	4	2	4	4
	3–6	15	10	10	14	14	13	13	4	2	2	2	1	0	2
	7–25	8	5	6	6	3	5	5	0	0	0	0	0	0	0
$T = 500$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$k = 25$	1	65	72	71	65	73	69	71	89	96	96	84	86	92	94
	2	14	10	13	15	12	13	11	8	3	3	11	10	7	4
	3–6	15	15	14	15	12	15	13	3	1	1	5	4	1	2
	7–25	6	3	2	5	3	3	5	0	0	0	0	0	0	0

TABLE II

Means of Order Selected and Frequency of Selecting the Same Order by the $FPEA_\alpha$ and $FPER_\alpha$ criteria

		$T=48$		$T=96$		$T=200$		$T=300$		$T=500$	
α	Statistic	$\beta=-0.8$	$\beta=0.3$	$\beta=-0.8$	$\beta=0.3$	$\beta=-0.8$	$\beta=0.3$	$\beta=-0.8$	$\beta=0.3$	$\beta=-0.8$	$\beta=0.3$
2	[a]	1.44	1.31	1.90	1.98	1.72	1.68	1.78	1.78	1.84	1.73
	[b]	1.45	1.03	1.81	1.71	1.81	1.87	2.13	2.11	2.07	2.02
	[c]	89	73	78	80	78	74	79	80	82	77
3	[a]	1.21	0.91	1.24	1.28	1.17	1.12	1.20	1.20	1.19	1.29
	[b]	1.17	0.74	1.29	1.24	1.36	1.41	1.26	1.24	1.38	1.39
	[c]	92	78	89	80	85	82	85	86	86	87
4	[a]	1.15	0.74	1.13	1.03	1.02	1.02	1.10	1.04	1.05	1.18
	[b]	1.05	0.65	1.17	1.06	1.09	1.13	1.14	1.10	1.19	1.22
	[c]	92	89	94	84	93	91	91	93	91	94

[a] Mean order selected by the $FPEA_\alpha$ criterion.
[b] Mean order selected by the $FPER_\alpha$ criterion.
[c] Frequency of selecting the same order.

especially when the fitted order s was greater than the true order h and/or when $\beta = -0.8$ and T is small.

For example, with $T = 48$ and $\beta = -0.8$, the iterative procedure failed to converge for all the fitted orders in 44 out of 100 simulations, i.e., for these 44 simulations no estimate of h was obtained. Although, for $T \geqslant 200$ an estimate of h was obtained in all 100 simulations for all the four values of β, the difficulties in obtaining convergence were still encountered, especially when $s > h$. For example, with $s = 10$, $T = 200$, and $\beta = -0.8$, the procedure failed to converge in 33 out of 100 simulations.

In a Ph.D. thesis submitted to Liverpool University, M. Mbago applies a criterion similar to (5.6), but with the moving average parameters estimated by an approximate maximum likelihood procedure suggested by Box and Jenkins [13] to simulated first, second, third, and fifth order processes. For the first order processes, his results are similar to that given in Table I for the AICM_α criterion. Again, no loss in efficiency results by using the FPEA_α, or the FPER_α, criteria; but estimation of h by using the Box–Jenkins approximate maximum likelihood procedure in conjunction with the AICM_α criterion is computationally more expensive.

It is instructive also to compare the relative performance of the FPEA_α and the FPER_α criteria. From Table I, it will be noticed that for each (T, β, α) configuration the frequency distributions of the order selected by these two criteria are similar, but the FPER_α criterion consistently selects the correct first order slightly more frequently than the FPEA_α criterion. A summary of the joint behaviour of these two criteria is given in Table II, where the frequency with which they select the same order is shown along with the average order selected. It is seen that the frequency with which these two criteria select the same order varies between 70% and 95%. For a fixed T this frequency increases as α increases, and vice versa.

The simulation results do not give specific guidance on what α to choose in practice. Reference may be made to Bhansali and Downham [12], Bhansali [9], Beenstock and Bhansali [8], Akaike [4], Smith and Spiegelhalter [34], and Atkinson [6] for a discussion of this question.

REFERENCES

[1] AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (B. N. Petrov and F. Csaki, Eds.), pp. 267–81. Akad. Kiadó, Budapest.

[2] AKAIKE, H. (1976). Spectrum estimation through parametric model fitting. In *Int. Union of Theoretical and Applied Mechanics: Symposium on Stochastic processes in Dynamics*, held at Southampton University, July 19–23, 1976.

[3] AKAIKE, H. (1978). Comments on "On model structure testing in system identification." *Internat. J. Control* **27** 323–324.

[4] AKAIKE, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika* **66** 237–242.
[5] ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.
[6] ATKINSON, A. C. (1980). A note on the generalized information criterion for choice of a model. *Biometrika* **67** 413–418.
[7] BARNDORFF-NIELSEN, O. AND SCHOU, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *J. Multivar. Anal.* **5** 408–419.
[8] BEENSTOCK, M. AND BHANSALI, R. J. (1980). Analysis of cocoa price series by autoregressive model fitting techniques. *J. Agric. Econom.* **31** 237–242.
[9] BHANSALI, R. J. (1979). A mixed spectrum analysis of the Lynx data. *J. Roy. Statist. Soc. Ser. A* **A142** 199–209.
[10] BHANSALI, R. J. (1980). Autoregressive and window estimates of the inverse correlation function. *Biometrika* **67** 551–566.
[11] BHANSALI, R. J. (1982). The estimation of the order of a moving average model from autoregressive and window estimates of the inverse correlation function. Research Report, University of Liverpool.
[12] BHANSALI, R. J. AND DOWNHAM, D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE-criterion. *Biometrika* **64** 547–551.
[13] BOX, G. E. P. AND JENKINS, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco.
[14] CHATFIELD, C. (1979). Inverse autocorrelations. *J. R. Statist. Soc.* **A142** 363–377.
[15] CLEVELAND, W. S. (1972). The inverse auto-correlations of a time series and their applications. *Technometrics* **14** 277–293.
[16] DURBIN, J. (1959). *Efficient estimation of parameters in moving average models*. *Biometrika* **28** 233–244.
[17] GRENANDER, U. AND ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.
[18] HANNAN, E. J. (1969). The estimation of the mixed moving average autoregressive systems. *Biometrika* **56** 579–593.
[19] HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.
[20] HANNAN, E. J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8** 1071–1081.
[21] HIPEL, K. W., MCLEOD, A. I., AND LENNOX, W. C. (1977). Advances in Box–Jenkins modelling 1. Model construction. *Water Resources Research* **13** 567–575.
[22] MCCLAVE, J. T. (1973). On the bias of autoregressive approximation to moving averages. *Biometrika* **60** 599–605.
[23] MCCLAVE, J. T. (1974). A comparison of moving average estimation procedures. *Comm. Statist.* **3** (9) 865–883.
[24] MCCLAVE, J. T. (1978). Estimating the order of moving average models: the max χ^2 method. *Comm. Statist.* **A7** (3) 259–276.
[25] NICHOLLS, D. F. (1972). On Hannan's estimation of ARMA models. *Austral. J. Statist.* **3** 262–269.
[26] OSBORN, D. R. (1976). Maximum likelihood estimation of moving average processes. *Ann. Eco. Soc. Meas.* **5** 75–87.
[27] PARZEN, E. (1974). Some recent advances in time series modelling. *IEEE Trans. Auto. Control.* **AC19** 723–730.
[28] QUENOUILLE, M. H. (1947). A large-sample test for the goodness of fit of autoregressive schemes. *J. R. Statist. Soc.*, **A110** 123–129.
[29] RAMSEY, F. L. (1974). Characterization of the partial auto-correlation function. *Ann. Statist.* **2** 1296–1301.

[30] RAO, C. R. (1975). *Linear Statistical Inferences and Its Applications.* Wiley, New York.
[31] RUDIN, W. (1974). *Real and Complex Analysis.* McGraw–Hill, New York.
[32] SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.
[33] SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126.
[34] SMITH, A. F. M. AND SPIEGELHALTER, D. J. (1980). Bayes factors and choice criteria for linear models. *J. R. Statist. Soc.*, **B42**, 213–220.
[35] SZEGÖ, G. (1939). *Orthogonal Polynomials.* Amer. Math. Soc. New York.
[36] WHITTLE, P. (1963). *Prediction and Regulation by Linear Least Square Methods.* English Univ. Press, London.